AML, $n = 4$), SJCRH (childhood AML, $n = 5$), or the Children's Cancer Group (childhood AML, $n = 5$) leukemia banks. The samples were processed as described (*13*), with the exception of the samples from SJCRH, which used a very different protocol. The SJCRH samples were subjected to hypotonic lysis (rather than Ficoll sedimentation), and RNA was prepared by an aqueous extraction (Qiagen).

24. Although the number of genes used had no significant effect on the outcome in this case (median PS for cross-validation ranged from 0.81 to 0.68 over a range of predictors using 10 to 200 genes, all with 0% error), it may matter in other instances. One approach is to vary the number of genes used, select the number that maximizes the accuracy rate in cross-validation, and then use the resulting model on the independent data set. In any case, we recommend using at least 10 genes for two reasons. Class predictors using a small number of genes may depend too heavily on any one gene and can produce spuriously high prediction strengths (because a large "margin of victory" can occur by chance due to statistical fluctuation resulting from a small number of genes). In general, we also considered the 99% confidence line in neighborhood analysis to be the upper bound for gene selection.
25. P. A. Dinndorf *et al.*, *Med. Pediatr. Oncol.* 20, 192 (1992); P. S. Master, S. J. Richards, J. Kendall, B. E. Roberts, C. S. Scott, *Blut* 59, 221 (1989); V. Buccheri *et al.*, *Blood* 82, 853 (1993).
26. M. Konopleva *et al.*, *Blood* 93, 1668 (1999).
27. A. W. Crawford and M. C. Beckerle, *J. Biol. Chem.* 266, 5847 (1991).
28. W. Ross, T. Rowe, B. Glisson, J. Yalowich, L. Liu, *Cancer Res.* 44, 5857 (1984).
29. Treatment failure was defined as failure to achieve a complete remission after a standard induction regimen including 3 days of anthracycline and 7 days of cytarabine. Treatment successes were defined as patients in continuous complete remission for a minimum of 3 years. FAB subclass M3 patients were excluded but samples were otherwise not selected with regard to FAB criteria.
30. J. Borrow *et al.*, *Nature Genet.* 12, 159 (1996); T. Nakamura, *et al.*, *ibid.*, p. 154; S. Y. Huang *et al.*, *Br. J. Haematol.* 96, 682 (1997).
31. E. Kroon *et al.*, *EMBO J.* 17, 3714 (1998).
32. P. Tamayo *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 96, 2907 (1999).
33. The SOM was constructed using our GENECLUSTER software (*32*), with a variation filter excluding genes with less than fivefold variation across the collection of samples.
34. For testing putative clusters derived from the SOM or chosen at random, we constructed class predictors with various number of genes (ranging from 10 to 200) and selected the one with the highest cross-validation accuracy rate (in this case, 20 genes).
35. A related approach would be to represent each cluster only as the subset of points lying near the centroid of the cluster.
36. Various statistical methods can be used to compare the predictors derived from the SOM-derived clusters with predictors derived from random classes. We compared the median prediction strength. Specifically, 100 predictors corresponding to random classes of comparable size were constructed, and the median PS for each predictor was determined. The performance for the actual predictor was then compared to the distribution of these 100 median PSs, to obtain empirical significance levels. The observed median PS in the initial data set was 0.86, which exceeded the median PS for all 100 random predictors; the empirical significance level was thus <1%. The observed median PS for the independent data set was 0.67, which exceed the median PS for all but 4 of the 100 random permutations; the empirical significance level was thus 4%.
37. Various approaches can be used to test classes C_1, C_2, ..., C_n arising from a multinode SOM. One can construct predictors to distinguish each pair of classes (C_i versus C_j) or to distinguish each class for the complement of the class (C_i versus not C_i). Here we used the pair-wise approach (C_i versus C_j). For cross-validation one can restrict attention to samples known to lie in the union of C_i and C_j. For an independent data set, one must examine all samples (because it is unknown which samples lie in the union of C_i and C_j). It may be possible to improve the statistical power of this test by using techniques for multiclass prediction.
38. Thirty-three ALL samples were tested by cross-validation using a 50-gene predictor. Thirty-two of 33 samples were correctly assigned as T-ALL or B-ALL; the remaining sample received a PS < 0.3, and no prediction was therefore made. Details are provided on our Web site.
39. T. R. Golub, unpublished results.
40. S. Turc-Carel *et al.*, *Cancer Genet. Cytogenet.* 19, 361 (1986); E. C. Douglass *et al.*, *ibid.* 45, 148 (1987).
41. We are grateful to S. Sallan, J. Ritz, K. Loughlin, S. Shurtleff, P. Kouzias, F. Smith, the Cancer and Leukemia Group B, and Children's Cancer Group for providing valuable patient samples. We thank R. Klausner, D. G. Gilliland, D. Nathan, G. Daley, J. Staunton, M. Angelo, A. Lal Ianc, P. Lee, Z. Kikinis, G. Acton, and members of the Lander and Golub laboratories for helpful discussions. This work was supported in part by the Leukemia Society of America (T.R.G); the National Institutes of Health and the Leukemia Clinical Research Foundation (C.D.B); and Affymetrix, Millennium Pharmaceuticals, and Bristol-Myers Squibb (E.S.L).

27 May 1999, accepted 17 September 1999

Sequencing Complex Polysaccharides

Ganesh Venkataraman,[1] Zachary Shriver,[2] Rahul Raman,[2] Ram Sasisekharan[2]*

Although rapid sequencing of polynucleotides and polypeptides has become commonplace, it has not been possible to rapidly sequence femto- to picomole amounts of tissue-derived complex polysaccharides. Heparin-like glycosaminoglycans (HLGAGs) were readily sequenced by a combination of matrix-assisted laser desorption ionization mass spectrometry and a notation system for representation of polysaccharide sequences. This will enable identification of sequences that are critical to HLGAG biological activities in anticoagulation, cell growth, and differentiation.

The chemical heterogeneity of polysaccharides, their structural complexity, and the lack of effective tools and methods have seriously limited the development of a sequencing approach that is rapid and practical, like that used for polynucleotides and polypeptides. This limitation is especially relevant in the study of glycosaminoglycan (GAG) complex polysaccharides, which are present at the cell surface and in the extracellular matrix (*1*, *2*). Heparin or heparan sulfate–like glycosaminoglycans (HLGAGs), a subset of GAGs, are currently used clinically as anticoagulants, and this function of HLGAGs has been assigned to a specific pentasaccharide sequence that is responsible for binding to antithrombin III (*3*). Recent progress in developmental biology, genetics, and other fields has resulted in a virtual explosion in the discovery of important roles for HLGAGs in the biological activity of morphogens (*4*) (for example, Wingless, Decapentaplegic, and Hedgehog), growth factors, cytokines, and chemokines (*5*); enzymes (*1*, *6*); and surface proteins of microorganisms (*7*). Although it is increasingly recognized that a specific sequence, typically from a tetra- to a decasaccharide in size, is responsible for HLGAGs' modulation of biological activity, in only a few cases is there any structural information regarding sequences (*8*). Therefore, accelerating our understanding of structure-function relationships for HLGAGs requires the development of rapid yet thorough sequencing methodologies.

There are many issues that have limited the development of sequencing techniques for HLGAGs. HLGAGs are chemically complex and heterogeneous, because the HLGAG chain can vary in terms of the number of disaccharide repeat units and possesses, within the disaccharide repeat unit, four potential sites for chemical modification. The basic disaccharide repeat unit of HLGAG is a uronic acid [α-L-iduronic acid (I) or β-D-glucuronic acid (G)] linked 1,4 to α-D-hexosamine (H) (Fig. 1A). Together, the four different modifications ($2^4 = 16$) for an I or G uronic acid isomer containing disaccharide give rise to $16 \times 2 = 32$ different plausible disaccharide units for HLGAGs. In contrast, four bases make up DNA, and 20 amino acids make up proteins. With these 32 building blocks, an octasaccharide could have over a million possible sequences, thereby making HLGAGs not only the most acidic but also the most information-dense biopolymers found in nature. There are no methods available to amplify or produce HLGAGs in large amounts, unlike the techniques that are available for DNA or proteins.

To handle the enormous information den-

[1]Harvard-MIT Division of Health Sciences and Technology, [2]Division of Bioengineering and Environmental Health, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

*To whom correspondence should be addressed. E-mail: ramnat@mit.edu

sity of HLGAGs, we have developed a hexadecimal notation system (base 16) to represent the four potential sites for chemical modification in the disaccharide repeat unit (*9*). The functional groups modified by sulfates can be represented by a binary digit (0 for no sulfate, and 1 for the presence of a sulfate). Thus, each possible HLGAG disaccharide can be assigned a one-letter code based on its hexadecimal value (Fig. 1B) (*10*). The isomer of the uronic acid can be assigned a positive (I) or a negative (G) sign with the same one-letter code. With such a coding scheme, mathematical operations (addition, subtraction, and so on) can be used to encode information such as susceptibility to enzymatic or chemical cleavage of specific sites and chemical properties that could facilitate sequence comparison or alignment (*9*). Furthermore, this code can be easily expanded in the future to add other modifications that might be observed, through the addition of binary digits to expand to a higher numerical base.

Different combinations of the 32 building blocks yield tetra-, hexa-, or longer saccharides, and it is possible to generate a list of theoretical molecular masses of all possible saccharide sequences as well as to uniquely assign the length of the saccharide based on the molecular weight (Fig. 1C). For instance, the minimum difference between any disaccharide and any tetrasaccharide is 101 daltons. Further, within an oligosaccharide larger than a disaccharide, there is a minimum difference in mass of 4.02 daltons that is accounted for by two acetate groups (84.08 daltons) replacing a sulfate group (80.06 daltons). Therefore, the length and the number of sulfates and acetates are readily assigned for a given oligosaccharide up to a tetradecasaccharide. The matrix-assisted laser desorption ionization mass spectrometry (MALDI-MS) technique described here enables the determination of the mass of HLGAG complex oligosaccharides (from di- to decasaccharides) to an accuracy of <1 dalton, with a sensitivity down to 100 fmol of material (*11*). The combination of the notation system and the accuracy of MALDI-MS enabled us to determine mass-identity relationships.

Thus, the overall strategy for sequencing HLGAGs essentially involves the reduction in size of the starting oligosaccharide into smaller fragments through the successive use of experimental constraints such as chemical and enzymatic degradation, as well as the use of MALDI-MS to determine the length of the saccharide and the number of sulfates and acetates in it. The elimination of sequences that do not satisfy the experimental constraints rapidly enables convergence to a unique sequence after successive iterations. Specific examples of how one would sequence HLGAGs with this process are illustrated below.

Example 1: MALDI-MS of the saccharide points to a mass of 1230.2 daltons, corresponding to an octasaccharide with 11 sulfates and no acetates (Fig. 2A). Analysis of this saccharide by capillary electrophoresis (CE) shows the presence of a trisulfated disaccharide ($\Delta U_{2S}H_{NS,6S}$; =D) and a disulfated disaccharide ($\Delta UH_{NS,6S}$; =5) with relative abundance of 3.1 (Fig. 2A, inset). Because both the I- and G-containing octasaccharides need to be considered (*12*), the number of possible combinations of sequences having these disaccharide units is 32 (Fig. 2D, part i). If heparinase I digestion data (Fig. 2B) are used as a constraint, only 4 out of the 32 sequences can result in fragments that correspond to these observed masses (Fig. 2D, part ii). Thus the heparinase I data fix two disaccharide units at the nonreducing end. However, for the reducing end, having +5 ($IH_{NS,6S}$) or −5 ($GH_{NS,6S}$) at the terminus or having +5 or −5 at the second position from the terminus are possible (Fig. 2D, part ii). To converge further on the octasaccharide sequence, we used heparinase III digestion data as a constraint (Fig. 2C). Of the sequences in Fig. 2D, part ii, the only sequence that can satisfy the observed masses of the products of heparinase III digestion is ±DDD-5, which corresponds to $\Delta U_{2S}H_{NS,6S}I_{2S}H_{NS,6S}I_{2S}H_{NS,6S}GH_{NS,6S}$.

Example 2: Here we illustrate the use of MALDI-MS and exoenzymes to sequence a nitrous acid–derived hexasaccharide. Partial digestion of the hexasaccharide with nitrous acid results in a ladder of hexa-, tetra-, and disaccharides (Fig. 3A). The nonreducing end of all the saccharides in the ladder can be sequenced simultaneously with exoenzymes (*13*). Upon treatment of the nitrous derived products with the exoenzymes (iduronate 2-O sulfatase, duronidase, and glucosamine 6-O sulfatase), the hexasaccharide and tetrasaccharide peaks were correspondingly reduced in mass, reflecting the removal of 2-O-sulfate, iduronate, and 6-O sulfate from the reducing end of the products (Fig. 3B). Com-

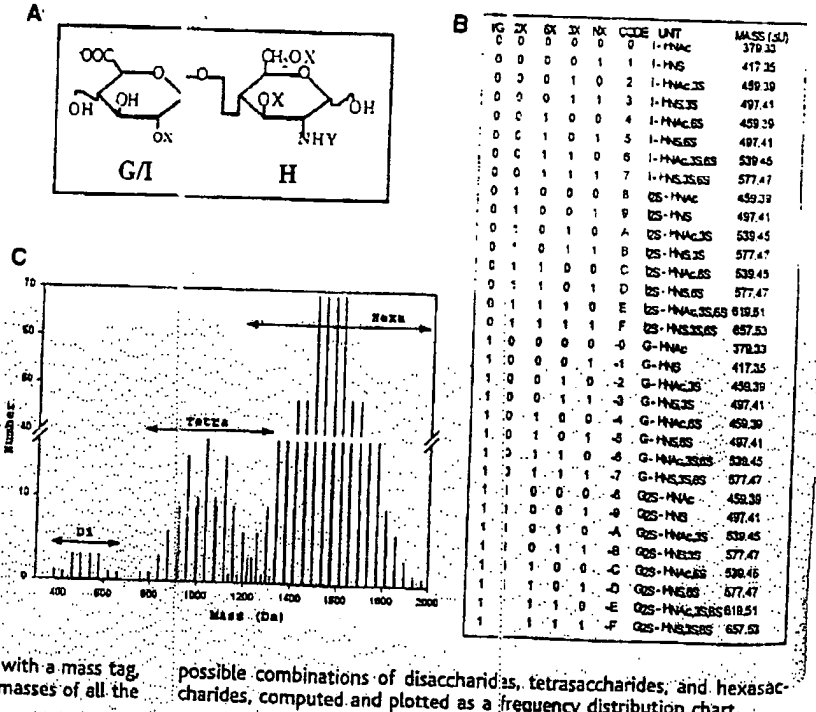


Fig. 1. HLGAG disaccharide repeating unit. (A) Disaccharide repeating unit containing a uronic acid [α-L-iduronate (I) or β-D-glucuronate (G)] with a 1-4 linkage to α-D hexosamine (H). The functional groups can be modified (primarily sulfation) at the 2-O, 3-O, 6-O (X), and N sulfation or acetylation (Y). (B) The hexadecimal notation system representing the U-H disaccharide units of HLGAGs. Columns 1 through 4 represent binary digits that indicate the presence (1) or absence (0) of sulfates at the 2, 6, 3, and N positions of the disaccharide repeat unit. A 0 at the N position represents a N-acetylated glucosamine. Column 5 shows the hexadecimal letter code that is encoded by the binary digits of columns 1 through 4. The hexadecimal units are assigned a sign based on the isomeric state of the uronic acid; positive (+) hexadecimal digits represent I-containing units and negative (−) hexadecimal digits represent G-containing units. The conventional chemical nomenclature is shown in column 6. The calculated molecular weight of the disaccharide units (as if they were internal in a chain) is shown in column 7. Free amine–containing disaccharides were not used in this study, but this form can be easily extended to include such modifications. Here and in Figs. 2 through 4, for chemical or enzymatic modifications to these disaccharides, the following symbols are used: 5-membered anhydromannitol ring, '; uronic acid with a C4-C5 unsaturated linkage (ΔU), ±; reducing end disaccharide unit with a mass tag, ⁺; disaccharide unit without the uronic acid, *. (C) The masses of all the possible combinations of disaccharides, tetrasaccharides, and hexasaccharides, computed and plotted as a frequency distribution chart.

IG	2X	6X	3X	NX	CODE	UNIT	MASS (Da)
0	0	0	0	0	0	I-HNAc	379.33
0	0	0	0	1	1	I-HNS	417.35
0	0	0	1	0	2	I-HNAc3S	459.39
0	0	0	1	1	3	I-HNS3S	497.41
0	0	1	0	0	4	I-HNAc6S	459.39
0	0	1	0	1	5	I-HNS6S	497.41
0	0	1	1	0	6	I-HNAc3S6S	539.45
0	0	1	1	1	7	I-HNS3S6S	577.47
0	1	0	0	0	8	I2S-HNAc	459.39
0	1	0	0	1	9	I2S-HNS	497.41
0	1	0	1	0	A	I2S-HNAc3S	539.45
0	1	0	1	1	B	I2S-HNS3S	577.47
0	1	1	0	0	C	I2S-HNAc6S	539.45
0	1	1	0	1	D	I2S-HNS6S	577.47
0	1	1	1	0	E	I2S-HNAc3S6S	619.51
0	1	1	1	1	F	I2S-HNS3S6S	657.53
1	0	0	0	0	-0	G-HNAc	379.33
1	0	0	0	1	-1	G-HNS	417.35
1	0	0	1	0	-2	G-HNAc3S	459.39
1	0	0	1	1	-3	G-HNS3S	497.41
1	0	1	0	0	-4	G-HNAc6S	459.39
1	0	1	0	1	-5	G-HNS6S	497.41
1	0	1	1	0	-6	G-HNAc3S6S	539.45
1	0	1	1	1	-7	G-HNS3S6S	577.47
1	1	0	0	0	-8	G2S-HNAc	459.39
1	1	0	0	1	-9	G2S-HNS	497.41
1	1	0	1	0	-A	G2S-HNAc3S	539.45
1	1	0	1	1	-B	G2S-HNS3S	577.47
1	1	1	0	0	-C	G2S-HNAc6S	539.45
1	1	1	0	1	-D	G2S-HNS6S	577.47
1	1	1	1	0	-E	G2S-HNAc3S6S	619.51
1	1	1	1	1	-F	G2S-HNS3S6S	657.53

sitional analysis by CE (by means of heinase I) indicated that disaccharide present the reducing end was $I_{2S}Man_{6S}$ (+D'). gether this information allows us to conrge to the sequence +DDD' ($I_{2S}H_{NS,6S}$ $H_{NS,6S}I_{2S}Man_{6S}$).

Example 3: MALDI-MS of a basic fibroast growth factor (FGF-2) binding saccha-le indicated a species with a mass of 2808 ltons, which corresponds to a tetradeca sulted decasaccharide with anhydromanitol at e reducing end of the saccharide (Fig. 4A, set). Compositional analysis of this sample sulted in two peaks corresponding to ±D $\Delta U_{2S}H_{NS,6S}$) and ±D' ($\Delta U_{2S}Man_{6S}$) in the tio 3:1. As this decasaccharide was derived nitrous acid degradation of heparin, the onic acid at the nonreducing end was not bserved by CE (232 nm). Therefore, the on-educing end residue was identified rough sequencing with exoenzymes to be D ($I_{2S}H_{NS,6S}$). The number of possible seuences with this composition is 16 (Fig. 4B, art i). Of the 16 sequences, those that could esult in the observed fragments upon hepainase I digestion of the decasaccharide (Fig. 4A) are shown in Fig. 4B, part ii. To resolve the isomeric state of the internal uronic acid +D versus −D, exhaustive digestion of the saccharide with heparinase I and heparinase III was performed. Heparinase I exhaustive digestion of the saccharide resulted in only two species corresponding to a trisulfated disaccharide (±D) and its anhydromannitol derivative, whereas heparinase III did not cleave the decasaccharide at all. The above facts taken together confirm that the sequence of the FGF-2 binding decasaccharide was DDDDD' [$(I_{2S}H_{NS,6S})_4I_{2S}Man_{6S}$].

Example 4: This is an example of the determination of a complex sequence. Compositional analysis of an AT-III binding saccharide indicated the presence of three building blocks, corresponding to $\Delta U_{2S}H_{NS,6S}$ (±D), $\Delta UH_{NAc,6S}$ (±4), and $\Delta UH_{NS,3S,6S}$ (±7) in the relative ratio of 2:1:1, respectively. The shortest oligosaccharide that can be formed with this composition corresponds to a decasaccharide, which is consistent with the MALDI-MS data. The total number of possible combinations of this tridecasulfated, singly acetylated, decasaccharide sequence with the above disaccharide building blocks is 320 (Fig. 5D, part i). Digestion of this decasaccharide with heparinase I resulted in four fragments (Fig. 5A). Of the 320 possible sequences, only 52 sequences satisfied heparinase I digestion data (Fig. 5D, part ii). The mass spectrum of the exhaustive digestion of the decasaccharide with heparinase I showed mass-to-charge ratio (*m/z*) values that corresponded to a trisulfated disaccharide and an octasulfated hexasaccharide, thereby further reducing the list of 52 sequences to 28 sequences (Fig. 5D, part ii). To further converge on the sequence, we used a "mass tag" at the reducing end of the saccharide, which enabled the identification of the saccharide sequence close to and at the reducing end (*11*). Treatment of the semicarbazide-tagged decasaccharide with heparinase II resulted in fragments with *m/z* values of 4805.0, 5264.6, 5320.9, 5381.7, 5897.7, and 5958.4 (Fig. 5B). The *m/z* values of 5320.9 and 5897.7 corresponded to a tagged tetrasulfated tetrasaccharide and to a tagged heptasulfated hexasaccharide, both containing the *N*-acetyl glucosamine residue. This places +4 or −4 ($UGH_{NAc,6S}$) at

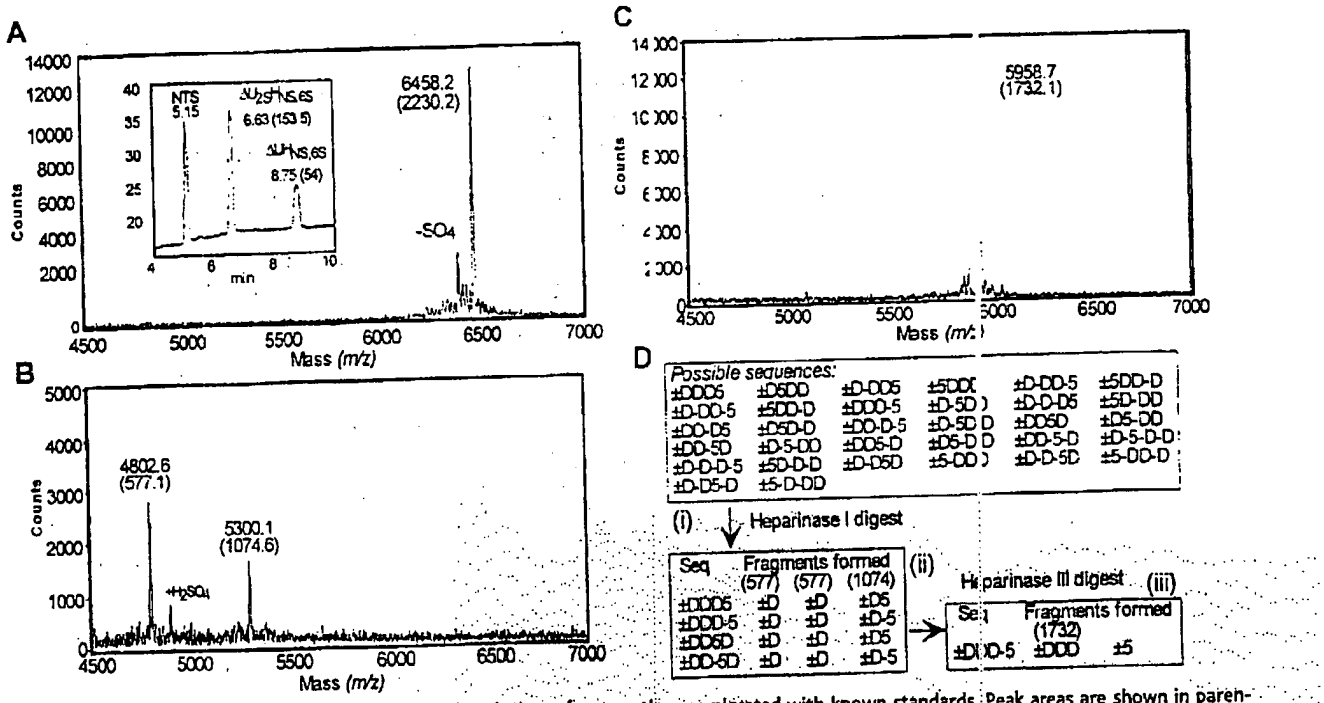


Fig. 2. Sequencing of octasaccharide. (A) A 10 μM solution of an octasaccharide sample was analyzed by MALDI-MS. The noncovalent complex of the peptide and the saccharide (*11*) was observed with a measured *m/z* value of 6458.2. Shown in parenthesis is the mass of the saccharide after subtraction of the peptide mass. $-SO_3$ indicates the loss of a labile sulfate of the octasaccharide substrate. Under relatively high laser intensities, larger oligosaccharides (octa in length or longer) can lose a sulfate. This undersulfated species is not a contaminant, as verified by CE. (Inset) Compositional analysis of the octasaccharide was completed by exhaustive digestion with heparinases I, II, and III and by analysis of the disaccharides with CE as described (*11*). Naphthalene trisulfonic acid (NTS; 1 μM) was run as an internal standard. Assignments of ΔU_{2S}-$H_{NS,6S}$ and ΔU-$H_{NS,6S}$ were made on the basis of the fact that they comigrated with known standards. Peak areas are shown in parentheses. (B) Digestion with heparinase I (*18*) resulted in a pentasulfated tetrasaccharide of *m/z* 5300.1 and a trisulfated disaccharide of *m/z* 4802.6. A H_2SO_4 adduct was also observed for the disaccharide as a result of matrix preparation conditions. (C) Digestion of octasaccharide with heparinase III (*18*) yielded a nine-sulfated hexasaccharide, *m/z* 5958.7. (D) shows the convergence of the octasaccharide. This octasaccharide contains a Δ4,5 uronic acid at the nonreducing end. (i) Master list of all possible octasaccharide sequences that can be obtained for three ±D and one ±5. The mass of the fragments resulting from digestion of octasaccharide with heparinase I is shown in (ii). Also shown in (ii) are sequences from (i) that can satisfy the heparinase I data. (iii) Sequence of octasaccharide from (ii) that satisfies the heparinase III digestion data (C).

the reducing end or one unit from the reducing end, thereby limiting the number of possible sequences from 28 to 6 (Fig. 5D, part iii). Partial nitrous acid digestion of the tagged as well as the untagged decasaccharide provided no additional constraints but confirmed the heparinase I data. Exhaustive nitrous acid digestion, however, gave only the reducing end tetrasaccharide (with and without the tag) as an unclipped product (Fig. 5C). This confirmed that +4 or −4 ($I/GH_{NAc,6S}$) was one unit away from the reducing end. Sequential use of exoenzymes uniquely resolved the isomeric state of the uronic acid as +4 and the reducing end disaccharide as −7, which is consistent with 4–7 being the key AT-III binding motif (*3*). Thus, we deduced the sequence of the AT-III binding decasaccharide as ±DDD4-7 ($\Delta U_{2S}H_{NS,6S}I_{2S}H_{NS,6S}I_{2S}H_{NS,6S}IH_{NAc,6S}GH_{NS,3S,6S}$) (*14*).

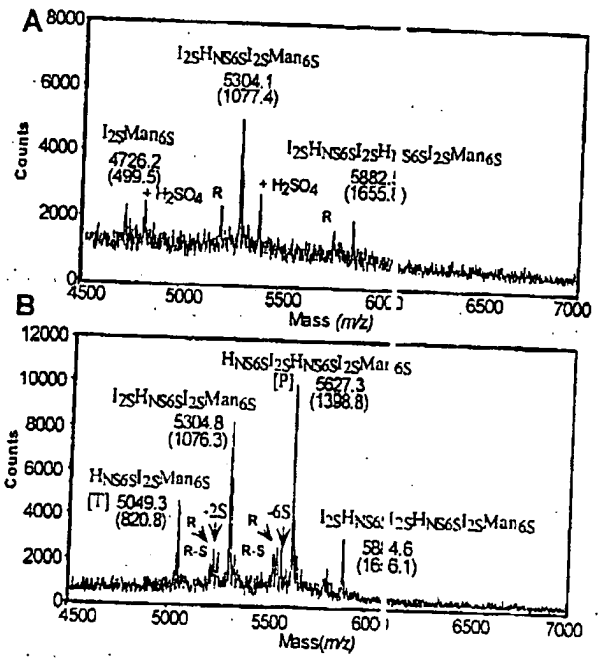


Fig. 3. Sequencing of hexasaccharide. (A) Nitrous acid treatment (20 min) of hexasaccharide (*18*) resulted in a ladder containing the starting material [with Man_{6S}] (*m/z* 5882.5), a pentasulfated tetrasaccharide [with Man_{6S}] (*m/z* 5304.1), and disulfated disaccharide (*m/z* 4726.2) [with Man_{6S}]. R indicates ring contraction (and a mass difference of 97 daltons) from a N-sulfated hexosamine to anhydromannose (*1*, *2*). Shown in parentheses is the mass of the saccharide after subtraction of the peptide mass. (B) The nitrous acid–treated hexasaccharide was subjected to iduronate 2-O sulfatase, iduronidase, and glucosamine 6-O sulfatase (*12*). The starting material (*m/z* 5884.6) and a pentasaccharide (P) (*m/z* 5627.3) corresponding to the removal of 2-O sulfate and an iduronate residue from hexasaccharide are visible. Also observed (marked as–6S) is the removal of the 6-O sulfate from P. For the tetrasaccharide (*m/z* 5304.8), the removal of 2-O sulfate (marked as–2S) and the iduronate residue resulted in a trisaccharide (T) (*m/z* 5049.3). The removal of the 6-O sulfate from T was observed only under exhaustive digestion conditions. The disaccharide was no longer detectable after incubation with the enzymes, because reduction of the charge on the disaccharide resulted in less efficient complexation with the basic peptide.

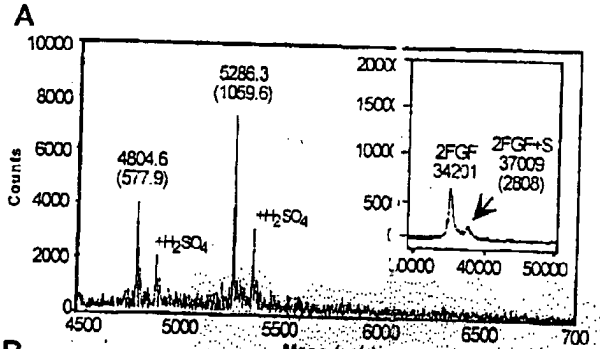
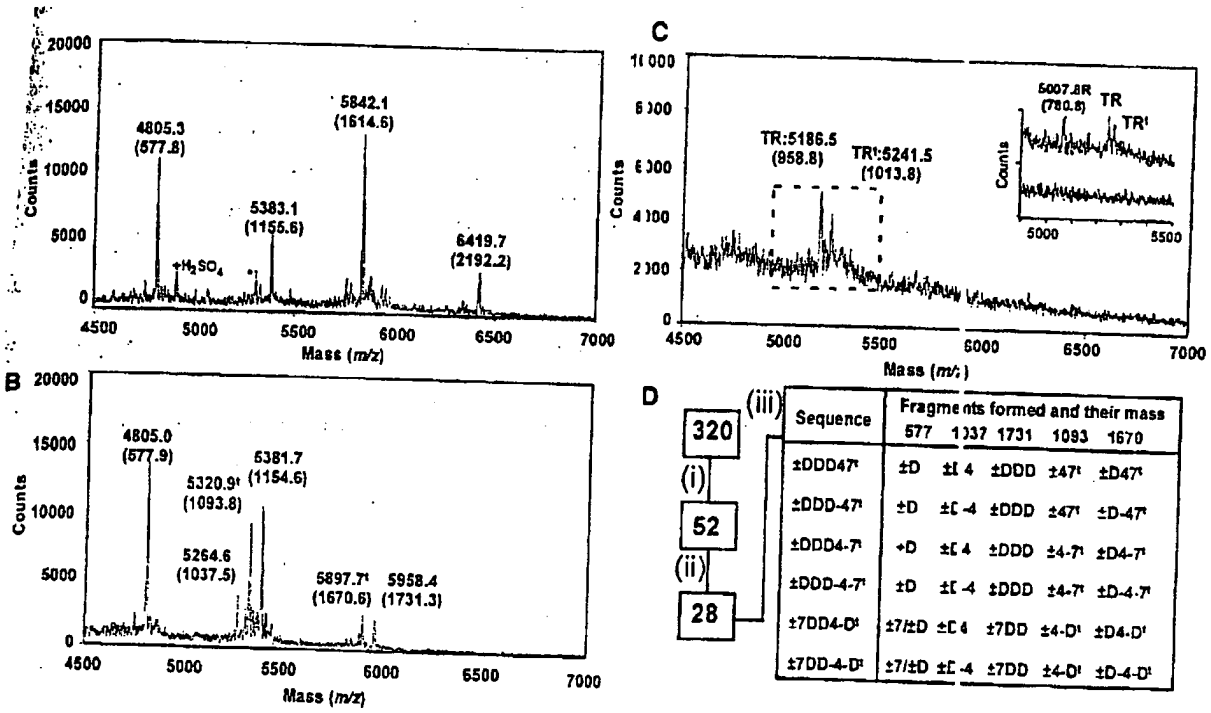


Fig. 4. Sequencing FGF-2 binding decasaccharide. (A) (Inset) MALDI-MS was performed to determine the mass and size of the saccharide as a complex with FGF-2 (*19*). Dimers of FGF-2 bound to the saccharide (S) yielding a species with a *m/z* of 37,009. By subtraction of FGF-2 molecular weight, the molecular mass of the saccharide was determined to be 2808, corresponding to a decasaccharide with 14 sulfates and an anhydromannitol at the reducing end. (A) Heparinase I digestion of the decasaccharide yielded a pentasulfated tetrasaccharide (*m/z* 5286.3) with an anhydromannitol at the reducing end and a trisulfated disaccharide of *m/z* 4804.6. (B) Convergence of the FGF binding decasaccharide sequence: the master list of 16 sequences derived from compositional analysis and exoenzyme sequencing of the nonreducing end. The disaccharide unit at the nonreducing end was assigned to be a +D using exoenzymes, and the anhydromannitol group at the reducing end is shown as '. The mass of the fragments resulting from digestion of decasaccharide with heparinase I are shown in (ii). Also shown in (ii) are the sequences from (i) that satisfy heparinase I digestion data. (iii) Sequence of decasaccharide from (ii) that satisfies the data from exhaustive digestion using heparinase I.

The sequencing approach can be readily extended to other complex polysaccharides by developing appropriate experimental constraints. For example, the dermatan/chondroitin mucopolysaccharides (DCMPs), consisting of a disaccharide repeat unit, are amenable to a hexadecimal coding system and MALDI-MS (*15*). As is observed for HLGAGs, in DCMPs a given mass has a unique signature associated with length and composition. For instance, the minimum difference between any disaccharide and any tetrasaccharide is 139.2 daltons; therefore, the length and the number of sulfates and acetates can be readily assigned for a given DCM oligosaccharide up to an octadecasaccharide. Similarly, in the case of polysialic acids (PSAs), which are present mostly as homopolymers of 5-*N*-acetylneuraminic acid (NAN) or 5-*N*-glycolylneuraminic acid (NGN), the hexadecimal coding system can be easily extended to NAN,NGN to encode the variations in the functional groups and enable a sequencing approach for PSAs (*16*).

The examples outlined here represent a practical and rapid sequencing methodology for HLGAGs in particular and for complex polysaccharides in general, so that it is possible to arrive at a single solution regardless of modifications of the polysaccharide chain. The advantages of the sequencing approach are threefold. First, through examination of all possible sequences for a oligosaccharide chain, it is possible to design further experimental constraints that will result in convergence to a single sequence in the fewest number of steps, thus making our sequencing approach rapid. The entire analysis in each example used only 1 pmol of material and took less than a day of experiments to arrive at the final sequence, as compared with the microgram-to-milligram amounts required for time-intensive methods (*13*). Second, because this approach starts from all possible sequences and eliminates those that do not conform to the experimental constraints, it ensures that unusual sequences are not eliminated with bias. Third, the method's flexibility not only allows for portability of the approach but will also facilitate development of a fully automated sequencing methodology (*17*).



. 5. Sequencing of AT-III binding decasaccharide. (A) Heparinase I est of the decasaccharide yielded four fragments. The major fragnts include a decasulfated, singly acetylated octasaccharide (*m/z* 9.7), a heptasulfated, singly acetylated hexasaccharide (*m/z* 5842.1), exasulfated tetrasaccharide (*m/z* of 5383.1), and a trisulfated disacride (*m/z* 4805.3). Also present is a contaminant (*), a pentasulfated asaccharide (*14*). The decasaccharide was first modified at the reducend to introduce a mass tag [Δ *m/z* of 56.1 is shown as t in (B) ough (D)]. Typical yields for the mass-tag labeling vary between 80 90%, as determined by CE. (B) Heparinase II digestion of the asaccharide gave the following products: *m/z* 5958.4 (a nine-sulfated asaccharide), *m/z* 5897.7 (tagged heptasulfated, singly acetylated asaccharide), *m/z* 5381.7 (hexasulfated tetrasaccharide), *m/z* 5320.9 (tagged tetrasulfated tetrasaccharide), *m/z* 5264.6 (tetrasulfated tetrasaccharide), and *m/z* 4805.0 (a trisulfated disaccharide). (C) Exhaustive nitrous acid treatment of decasaccharide essentially gives one tetrasulfated, singly acetylated, anhydromannitol tetrasaccharide species designated T (one tagged *m/z* 5241.5 and one untagged *m/z* 5186.5). R indicates ring contraction. As shown in the inset, treatment of this tetrasaccharide with iduronidase (and not glucuronidase) results in a species of *m/z* 5007.8, corresponding to the removal of iduronate residue. Further treatment with exoenzymes in the following order only (glucosamine 6-O sulfatase, hexosamidase, and glucuronidase) results in the complete digestion of the trisaccharide. (D) Convergence of the AT-III binding decasaccharide sequence from 3.0 possible sequences to 52 to 28 to 6 to the final sequence.

References and Notes

1. E. Conrad, *Heparin Binding Proteins* (Academic ress, San Diego, CA, 1998).
2. Ernst *et al.*, *Crit. Rev. Biochem. Mol. Biol.* **30**, 387 1995); A. Varki, *Glycobiology* **3**, 97 (1993); R. D. osenberg *et al.*, *J. Clin. Invest.* **99**, 2062 (1997).
3. I. Lindahl *et al.*, *Thrombosis Res.* **75**, 1 (1995); D. A. ane and U. Lindahl, *Heparin—Chemical and Biological Properties Clinical Applications* (CRC Press, Boca aton, FL, 1989).
4. Wodarz and R. Nusse, *Annu. Rev. Cell Dev. Biol.* **14**, 9 (1998); S. Cumberledge *et al.*, *Trends Genet.* **11**, 21 (1997).
5. R. Bernfield *et al.*, *Annu. Rev. Cell Biol.* **8**, 365 (1992); C. Rapraeger, *Curr. Opin. Cell Biol.* **5**, 844 (1993); B. Spom and A. B. Roberts, *Peptide Growth Factors id Their Receptors* (Springer-Verlag, Berlin, 1990); M. dmivitra *et al.*, *FASEB J.* **10**, 1270 (1996); L. Kjellén and Lindahl, *Annu. Rev. Biochem.* **60**, 443 (1991).
6. L. Jackson *et al.*, *Physiol. Rev.* **71**, 481 (1991).
7. L. Hung *et al.*, *Virology* **257**, 156 (1999); J. G. Joyce *al.*, *J. Biol. Chem.* **274**, 5810 (1999); A. Jacquet *et* , *Virus Res.* **53**, 197 (1998); C. McCormick *et al.*, *ture Genet.* **19**, 158 (1998); Y. Chen *et al.*, *Nature d.* **8**, 866 (1997).
8. Lindahl, *J. Biol. Chem.* **273**, 24979 (1998); R. D. senberg *et al.*, *J. Clin. Invest.* **99**, 2062 (1997); G. iksman and A. B. Herr, *Nature Struct. Biol.* **7**, 527 998).
9. hexadecimal system chosen here and shown in . 1B is based on a property-encoded notation tem (R. Raman *et al.*, in preparation).
10. Here, the disaccharide building blocks of HLGAGs are first represented with the uronic acid at the nonreducing end and the hexosamine at the reducing end (U-H). It is also possible to have H at the nonreducing end to define the repeating disaccharides in the HLGAG chain (H-U), although this will not conceptually change the analysis.
11. The methodology used for the analysis of acidic polysaccharides involves the detection of these saccharides as noncovalent complexes with a basic peptide or protein. Interactions between the polysaccharide and the protein or peptide allow for the ionization of normally labile HLGAGs as intact species and their detection in the positive ion mode. With the mass spectrometer in the linear mode, this technique is able to detect as little as 100 fmol of material [P. Juhasz and K. Biemann, *Carbohydr. Res.* **270**, 131 (1995); P. Juhasz and K. Biemann, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 4333 (1994); A. J. Rhomberg *et al.*, *ibid.* **95**, 4176 (1998); S. Ernst *et al.*, *ibid.*, p. 4182; A. J. Rhomberg *et al.*, *ibid.*, p. 12232].
12. Treatment with heparinases leaves behind a 4-5 double bond on the uronic acid, resulting in the loss of information regarding the isomeric state of the uronic acid. Heparinase I primarily clips I_S containing glycosidic linkages, whereas heparinase III clips primarily G containing glycosidic linkages. However, recent observations have noted that, under certain conditions, both heparinase I and III possess secondary substrate specificities that is, heparinase I can clip at G_{2S} in highly sulfated regions of HLGAGs, whereas heparinase III can clip at I in unsulfated regions of HLGAGs. Using defined substrates, we defined the cleavage conditions for heparinase I and III under our reaction conditions. We found that there were certain well-defined parameters under which heparinase I and III clip at their respective primary sites and not at their secondary sites (supplementary data is available at *Science* Online at www.sciencemag.org/feature/data/1042118.shl). In part, these studies defined the designated "short" and "exhaustive" reaction conditions outlined below (*18*). In each case, through the use of our sequencing strategy and independent experimental constraints (that is, incomplete nitrous acid followed by exoenzymes or exhaustive nitrous acid and compositional analysis), we can confirm the sequence assignment obtained with heparinases as experimental constraints.
13. Other methods including the use of exoenzymes are being developed for sequencing saccharides, and for the most part require sample isolation and repurification [J. H. Turnbull *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 2701 (1999), and references therein; R. Vives *et al.*, *Biochem. J.* **339**, 767 (1999)]. However, the MALDI-MS approach for sequencing saccharides with exoenzymes does not require repurification of the saccharide after manipulations. The MALDI mass spectrogram is able to generate all the sequence information of all the saccharide fragments in a "ladder" form in one spectrum.
14. The AT-III binding decasaccharide was a gift from R. Linhardt. The sequence of this decasaccharide has been reported to be D4-7DD, on the basis of NMR spectroscopy [Y. Toida *et al.*, *J. Biol. Chem.* **271**,

32040 (1996)]. Such a sequence should show the appearance of a tagged D or DD residue at the reducing end. However, we find all the different experiments used in the elucidation of the decasaccharide sequence to be consistent with each other in the appearance of a 4-7 tagged product and not a D (or a DD) product. This saccharide does not contain an intact AT-III binding site, as proposed. Therefore, we sought confirmation of our proposed sequence through the use of integral glycan sequencing (IGS), which agreed with our analysis.

15. DCMPs are found in dense connective tissues such as bone and cartilage. The basic repeat unit of DCMP can be represented as $-(\beta\ 1\rightarrow4)\ U_{2X}-(\alpha/\beta\ 1\rightarrow3)\ Gal_{NAc4X6X}-$, where U is uronic acid, Gal_{NAc} is a *N*-acetylated galactosamine, and there are 16 disaccharide building blocks for DCMP. Like the heparinases that degrade HLGAGs, there are distinct chondroitinases and other chemical methods available that clip at specific glycosidic linkages of DCMP and serve as experimental constraints. Furthermore, because DCMPs are acidic polysaccharides, the MALDI-MS techniques and methods used for HLGAGs can be readily extended to the DCMPs.

16. The monomeric units of NAN and NGN are linked by α 2-8 glycosidic linkages and can be modified at the 4-O, 7-O, and 9-O positions. Methods of purifying and characterizing PSA oligosaccharides using high-performance liquid chromatography, CE, and MS have very recently been established. In addition, chemical and exosialydases and endosialydases that cleave at distinct glycosidic linkage of PSA are available and can serve as experimental constraints.

17. The sequencing approach presently uses a brute force method because many of the rules regarding the specificity of enzymes that degrade and modify complex polysaccharides are in the developmental stage. Once these rules or constraints are fully developed, more intelligent algorithms such as a genetic algorithm or Monte Carlo optimization could be used to search a much narrower search space for the correct sequence. The combination of more efficient constraints and algorithms will thus lead to a fully automated sequencing approach.

18. The use of heparinases was essentially as described in references in (*11*). Digests were designated as either short or exhaustive. Short digests were completed with 50 nM enzyme for 10 min. Exhaustive digests were completed with 200 nM enzyme for either 4 hours or overnight. Partial nitrous acid cleavage was completed using a modification of published procedures. Briefly, to an aqueous solution of saccharide was added a 2× solution of sodium nitrite in HCl so that the concentration of nitrous acid was 2 mM and that of HCl was 20 mM. The reaction was allowed to proceed at room temperature with quenching of aliquots at various time points through the addition of 1/4 of 200 mM sodium acetate and of bovine serum albumin (1 mg/ml; pH 6.0). Exhaustive nitrous acid cleavage was completed by reacting saccharide with 1 M nitrous acid in HCl overnight at room temperature. In both cases, it was found that the products of nitrous acid cleavage could be sampled directly by MALDI without further cleanup and without the need to reduce the anhydromannose residues to anhydromannitol. The entire panel of HLGAG-degrading exoenzymes was purchased from Oxford Glycosystems (Wakefield, MA) and used as suggested by the manufacturer.

19. G. Venkataraman *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 1892 (1999).

20. Supported in part by the Massachusetts Institute of Technology and NIH (grant number GM 57073). We thank K. Biemann for help and input on the MALDI-MS; K. Pojasek for technical help with MS; R. Linhardt and J. J. Turnbull for saccharides and corroboration of our AT-III decasaccharide microsequencing results with exoenzyme sequencing (IGS); and J. Essigmann, A. Grodzinsky, R. Langer, and V. Sasisekharan for critical comments.

26 May 1999; accepted 1 September 1999

Stability and Variability in Competitive Communities

A. R. Ives,* K. Gross, J. L. Klug

Long-term variability in the abundance of populations depends on the sensitivity of species to environmental fluctuations and the amplification of environmental fluctuations by interactions among species. Although competitive interactions and species number may have diverse effects on variability measured at the individual species level, a combination of theoretical analyses shows that these factors have no effect on variability measured at the community level. Therefore, biodiversity may increase community stability by promoting diversity among species in their responses to environmental fluctuations, but increasing the number and strength of competitive interactions has little effect.

The stability of an ecological community is thought to depend on the number of species it contains and the strengths of interactions between them (*1*). Stability also depends on the organizational level at which it is measured. In model ecosystems, May (*2*) showed that increasing the strength and number of species interactions decreases the stability of individual species' dynamics. In contrast, experimental studies (*3*) showed that large communities of plant competitors may be more stable than small communities when stability is measured in terms of the total biomass of all species.

The stability of ecological communities is commonly characterized by one of five properties: mathematical stability, resilience, resistance, persistence of species, and variability (*4*). Of these, variability is the most frequently measured, yet least understood from a theoretical perspective. Our analyses address the effects of interspecific competition and species diversity on variability in the biomass of individual species and variability in total community biomass. Environmental fluctuations cause short-term (year-to-year) changes in species biomasses. Species interactions then act as a filter through which short-term environmental variability is translated into long-term variability in biomass. Here, we ask how this filter is affected by the strength of species interactions and the number of species in a community. We use an analytical approximation to derive general predictions, and explicit numerical models for quantitative illustration.

The general form of the systems we consider is

$$x_i(t+1) = x_i(t)F\left[x_i(t), \sum_{j\neq i}^{n}\alpha_{ij}x_j(t), \varepsilon_i(t), r_i, K_i\right] \quad (1)$$

where $x_i(t)$ is the biomass of species i in year t; $F[\ ]$ is a function giving the change in biomass between years; r_i and K_i are the species-specific intrinsic rate of increase and carrying capacity, respectively; and α_{ij} is the competition coefficient measuring the effect of species j on species i. The model is parameterized so that, in the absence of environmental variability, when $r_i > 1$, populations show overcompensating dynamics, and when $r_i < 1$ biomasses tend to approach equilibrium monotonically. Environmental fluctuations are included with species-specific random variables $\varepsilon_i(t)$. To account for similarities among species in their responses to environmental fluctuations, we assume that the $\varepsilon_i(t)$ values are correlated with correlation coefficient ρ.

We measure variability at two levels. For the species level, we use the sum of the variances of the biomasses of constituent species $V_n^s = \sum_{i=1}^{n} V[x_i(t)]$, where n is the number of species in the community. For the community level, we consider the variance of the aggregate biomass in the community, $V_n^c = V[\sum_{i=1}^{n} x_i(t)]$. Measuring variability in this way relates our results directly to experimental studies employing variances (*3–5*).

Analytical approaches provide general conclusions that are independent of detailed model assumptions. Although variances in species biomasses produced by Eq. 1 depend on nonlinearities in $F[\ ]$, variances can be approximated with the first-order autoregression model (*6*)

$$\mathbf{X}(t+1) = \mathbf{C}\mathbf{X}(t) + \mathbf{E}(t) \quad (2)$$

where $\mathbf{X}(t)$ is a vector of species biomasses $\Delta x_i(t)$ relative to the mean $[\Delta x_i(t) = x_i(t) - \bar{x}]$ and $\mathbf{E}(t)$ is a vector of random variables that contains the environmentally driven changes in species biomasses from one year to the

Department of Zoology, University of Wisconsin-Madison, Madison, WI 53706, USA.

*To whom correspondence should be addressed. E-mail: arives@facstaff.wisc.edu